

# DOS EJEMPLOS DE DISEÑOS EXPERIMENTALES SOBRE UN PROGRAMA DE PERFECCIONAMIENTO DE PROFESORES

Por GONZALO VÁZQUEZ GÓMEZ

## INTRODUCCIÓN

Los estudios que presento aquí están entresacados de un programa de perfeccionamiento para profesores titulado «Experiencias de investigación activa en el proceso educativo» realizado en el I.C.E. de la Universidad de Navarra en los años 1973 y 1974. Dicho programa tiene como uno de sus objetivos principales el de «capacitar a los profesores para que lleven a cabo la programación, el desarrollo y la evaluación de un proyecto de investigación activa».

Dentro de las actividades del programa, en estos dos años, se han llevado a cabo varios estudios centrados en las experiencias de aprendizaje realizadas por los propios participantes. Dado que una de las áreas de trabajo del programa es el estudio y práctica en el diseño y evaluación de proyectos «experimentales», se viene procurando realizar y mostrar diferentes trabajos sobre las mismas actividades de los participantes.

Traeré aquí dos ejemplos, inscribibles, uno en lo que unos

autores llaman diseño preexperimental o diseño evaluativo, y el otro en los diseños factoriales <sup>1</sup>.

### A) UN EJEMPLO DE DISEÑO PREEXPERIMENTAL

Se trata, realmente, de un diseño del tipo «*pretest-posttest* de un solo grupo» <sup>2</sup>.

1. *Objetivo*: mejorar —de forma estadísticamente significativa— la apreciación por los participantes en el programa de la corrección o incorrección en la formulación escrita de 50 objetivos educativos.
2. *Población implicada*: N = 19 (el conjunto total de los participantes).
3. *Variable cuyos efectos se trata de medir*: trabajo en grupos pequeños, durante veinte minutos, sobre el documento: DA - 131. «Cuestionario sobre las características de un objetivo concreto» <sup>3</sup>.
4. *Instrumento de medida* (a aplicar como *pretest* y como *posttest*): DA - 126. «Cuestionario de evaluación de objetivos» <sup>4</sup>.
5. *Tratamiento de los datos*: aplicación de la prueba de «t» para verificar la hipótesis nula respecto de la diferencia entre las medidas de las muestras.
6. *Datos obtenidos y análisis de los datos*.

<sup>1</sup> Cfr. F. N. KERLINGER, *Foundations of Behavioral Research*. London, 1973, 2nd ed., pág. 317; D. CAMPBELL y J. STANLEY, *Diseños experimentales y cuasiexperimentales en la investigación social*. B. Aires, 1973, págs. 19 y ss.; D. FOX, *The Research Process in Education*. N. York, 1970, pág. 453, y J. FERNÁNDEZ HUERTA, «Diseños factoriales complejos en la investigación psicodidáctica». REP. XIV, 54, abril-junio, 1956, págs. 172-83.

<sup>2</sup> La información que sigue está tomada de G. VÁZQUEZ GÓMEZ, *Ejemplo de diseño preexperimental*. Pamplona, ICE de la Universidad de Navarra. Abril 1974. Multicopiado.

<sup>3</sup> Se trata de un documento elaborado por mí y editado por el propio Instituto, abril 1973. Multicopiado.

<sup>4</sup> Idem, mismo lugar y fecha.

TABLA 1

PARTICIPANTE	«Pretest» (x)	«Postest» (Y)	Observaciones
A ... ..	48	47	5
B ... ..	43	45	
C ... ..	43	—	
D ... ..	36	42	
E ... ..	31	42	
F ... ..	31	39	6
G ... ..	30	33	
H ... ..	26	22	
I ... ..	24	40	
J ... ..	23	40	
K ... ..	22	37	
L ... ..	22	35	
M ... ..	22	27	
N ... ..	18	38	
O ... ..	17	36	
P ... ..	15	21	
Q ... ..	14	29	
R ... ..	14	26	
S ... ..	7	43	

$$N_x = 17$$

$$\bar{X}_x = 25,647$$

$$S_x = 9,554$$

$$\sigma_{\bar{x}} = 9,848$$

$$\sigma_{\bar{x}\bar{x}} = 2,389$$

$$N_y = 17$$

$$\bar{X}_y = 35,294$$

$$S_y = 7,614$$

$$\sigma_y = 7,848$$

$$\sigma_{\bar{x}y} = 1,903$$

$$\bar{X}_y - \bar{X}_x = 9,647$$

$$t = 3,158 (p < .005)$$

### Crítica del preexperimento

#### Argumentos a favor:

- el «experimento» se ha centrado sobre todo el conjunto de participantes. En tal sentido, no ha habido *selección* alguna;
- tampoco ha habido *mortalidad*. Propiamente hablando

<sup>5</sup> No se tuvo en cuenta su respuesta en el *pretest* porque, al contestar al *postest*, se equivocó transcribiendo mal los resultados (saltándose alguna celda, error que no se pudo descifrar con seguridad).

<sup>6</sup> También se excluyeron estos datos porque, según aclaración propia del mismo participante, no había entendido bien las instrucciones del *pretest*.

ha habido una «baja» a nivel de *posttest*, pero se ha debido al error —ya comentado— en la transcripción de los resultados, por lo que no está afectada por el sesgo de la voluntariedad y otra a nivel de *pretest*, también involuntaria.

*Argumentos en contra:*

Por no haber grupo de control, no se han podido controlar rigurosamente las siguientes fuentes de invalidación:

- *historia*: acontecimientos específicos ocurridos entre la primera y la segunda medición (aparte de la variable experimental).
- *maduración*: proceso interno de los participantes, que se opera al paso del tiempo.

No obstante, dado el poco tiempo trascurrido entre el *pretest* y el *posttest* cabe pensar que esta segunda fuente de invalidación haya sido poco importante. No así la primera, ya que ha mediado otro trabajo de lectura y discusión de un documento sobre el trabajo por objetivos.

Por haber sido *exactamente* la misma prueba la aplicada en ambas situaciones, cabe pensar que haya influido positivamente en el *posttest* el aprendizaje producido por la primera aplicación<sup>7</sup>.

*Evaluación del objetivo:*

Se ha logrado alcanzar esta diferencia significativa al nivel de confianza del 995 por 1.000.

B) EJEMPLOS DE DISEÑO FACTORIAL

Realizado con varios grupos experimentales y aplicando un *pretest* y un *posttest*.

<sup>7</sup> Para una crítica más precisa, cfr. D. CAMPBELL y J. STANLEY, *Op. et loc. cit.*

## 1. *Objetivos:*

- 1.1. Conseguir un incremento —estadísticamente significativo— en los conocimientos de estadística descriptiva y muestral básicos para realizar la evaluación de un proyecto de experimentación didáctica.
- 1.2. Conseguir un incremento en el interés por la aplicación de la estadística a la evaluación de un proyecto didáctico.
- 1.3. Probar la eficacia relativa de tres tipos de actividades para la adquisición o mejora de los conocimientos y actitudes hacia la estadística (según el esquema que cita más abajo; la hipótesis era que el método 3.º era superior al 2.º y éste al 1.º).

## 2. *Metodología:*

- 2.1. Introducción al trabajo posterior: presentación de la experiencia<sup>8</sup>.
- 2.2. Aplicación del *pretest* de conocimientos y actitud ante la estadística a partir del documento: DA - 132. «Prueba de evaluación de conocimientos de estadística»<sup>9</sup>.
- 2.3. Corrección de las respuestas dadas al *pretest*.
- 2.4. Composición, en función de los resultados anteriores, de los tres grupos experimentales homogéneos (A, B y C), adscribiendo a cada uno un conjunto de ocho participantes.
- 2.5. Realización del experimento didáctico según este esquema:

---

<sup>8</sup> Por exigencia del tiempo se previó la inversión de las fases 1.ª y 2.ª. Así se daría ocasión a corregir el *pretest* entre tanto se hacía la presentación de la experiencia.

<sup>9</sup> Cfr. G. VÁZQUEZ GÓMEZ y C. SALVADOR GARCÍA, *Prueba de evaluación de conocimientos de estadística*. Pamplona, ICE de la Universidad de Navarra, abril 1973. Multicopiado.



GRUPO	ACTIVIDADES
A ... ..	1. Clase magistral (seguida a través del C.C.T.V.) sobre el mismo tema de la prueba.
B ... ..	1. Lectura del documento DA - 134. «Fundamentos estadísticos para la evaluación de experiencias didácticas» <sup>10</sup> . 2. Clase magistral (por C.C.T.V.) sobre el tema.
C ... ..	1. Lectura del documento DA - 134. 2. Discusión sobre el tema, con planteamiento de dificultades por los participantes.

2.6. Aplicación del *postest* de conocimientos y actitud respecto de la estadística<sup>11</sup>.

### 3. Evaluación:

- 3.1. Aplicación previa (*pretest*) de la prueba DA - 132.
- 3.2. Aplicación posterior (*postest*) de la misma prueba DA - 132.
- 3.3. Verificación de las pruebas de «F» y de «t» para aceptar o rechazar la homogeneidad de las varianzas y de las medias de las distintas muestras.
- 3.4. Nivel de significación elegido:  $p < .05$ .

<sup>10</sup> De los mismos autores, lugar y fecha de edición.

<sup>11</sup> La pregunta acerca de la actitud estaba formulada como una cuestión de elección múltiple, donde representa la actitud más favorable y 5 la menos favorable. Téngase en cuenta esta escala para interpretar, más adelante, las diferencias entre las actitudes.

## 4. Análisis de los resultados:

## 4.1. A nivel de pretest.

Tal como estaba previsto, se aplicó la prueba previa de conocimientos y de actitud ante la estadística. Los resultados obtenidos son los que se incluyen en la tabla núm. 2, en la que se hace notar el grupo al que se adscribió a cada participante.

TABLA 2

DISTRIBUCION DE LAS PUNTUACIONES DE CONOCIMIENTOS Y DE ACTITUDES (PRETEST) CON INDICACION AL GRUPO AL QUE SE ADSCRIBIO CADA PARTICIPANTE

PARTICIPANTE	PUNTUACIONES EN EL «PRETEST»		Grupo al que se le adscribió
	Conocimientos	Actitud	
A ... ..	12	1	1.º
B ... ..	11	3	2.º
C ... ..	9	5	3.º
D ... ..	8	3	3.º
E ... ..	8	2	2.º
F ... ..	6	3	1.º
G ... ..	6	2	1.º
H ... ..	6	3	2.º
I ... ..	6	2	3.º
J ... ..	6	3	3.º
K ... ..	6	2	2.º
L ... ..	5	1	1.º
M ... ..	5	3	1.º
N ... ..	5	2	2.º
O ... ..	5	3	3.º
P ... ..	4	3	3.º
Q ... ..	4	2	2.º
R ... ..	3	5	1.º
S ... ..	3	2	1.º
T ... ..	3	4	2.º
U ... ..	3	3	3.º
V ... ..	3	2	3.º
X ... ..	2	2	2.º
Y ... ..	1	4	1.º

## Estadísticos y parámetros correspondientes a esta prueba:

	Conocimientos	Actitud
$N$ .....	24	24
$\bar{X}$ .....	6,125	2,708
$S$ .....	2,818	1,020
$S^2$ .....	7,942	1,040
$\sigma$ .....	2,879	1,042
$\sigma_{\bar{x}}$ .....	0,589	0,213

Como se puede observar en la tabla núm. 2, la adscripción de los participantes a los grupos se realizó utilizando como criterio preferente el nivel de sus conocimientos previos. No obstante, en el caso de empate se procuró repartirlos teniendo en cuenta también el criterio actitudinal. Así, los participantes F a K que lograron seis respuestas correctas se inscribieron en los grupos de esta forma:

Puntuación conocimiento-actitud	Participantes en los grupos		
	1	2	3
6 - 2 —————	G	K	I
6 - 3 —————	F	H	J

Con posterioridad a la formación de los grupos, nos interesamos también en el cálculo de la significación de las diferencias entre las varianzas y las medidas de las muestras de hombres y de mujeres, tanto para los conocimientos como respecto de las actitudes. Las conclusiones son éstas:

— prueba de conocimientos:

• mujeres: $\bar{X} = 3,000$	hombres: $\bar{X} = 2,000$
$\sigma = 0,970$	$\sigma = 0,756$



Análisis de la homogeneidad de la varianza (prueba de Barlett):

$$F = 1,109 \ (p < .05)$$

— prueba de actitudes:

$$\begin{array}{ll} \bullet \text{ mujeres: } \bar{X} = 5,824 & \text{hombres: } \bar{X} = 6,857 \\ \sigma = 3,044 & \sigma = 2,662 \end{array}$$

Análisis de la homogeneidad de la varianza (prueba de Barlett):

$$F = 0,169 \ (p > .05)$$

Para comprobar la eficacia de los criterios de composición de los grupos experimentales hemos procedido a realizar el correspondiente análisis de la homogeneidad, tanto respecto de los conocimientos como de la actitud previa

TABLA 4

DISTRIBUCION DE LAS PUNTUACIONES DE CONOCIMIENTOS  
PREVIOS EN LOS TRES GRUPOS

GRUPO 1		GRUPO 2		GRUPO 3	
Participante	Puntuación	Participante	Puntuación	Participante	Puntuación
A ... ..	12	B ... ..	11	C ... ..	9
F ... ..	6	E ... ..	8	D ... ..	8
G ... ..	6	H ... ..	6	I ... ..	6
L ... ..	5	K ... ..	6	J ... ..	6
M ... ..	5	N ... ..	5	O ... ..	5
R ... ..	3	Q ... ..	4	P ... ..	4
S ... ..	3	T ... ..	3	U ... ..	3
Y ... ..	1	X ... ..	2	V ... ..	3

## ESTADÍSTICOS Y PARÁMETROS

N	...	8	8	8
$\bar{X}$	...	5,125	5,625	5,500
S	...	3,059	2,690	2,062
S <sup>2</sup>	...	9,359	7,234	4,250
$\sigma$	...	3,270	2,875	2,204
$\sigma\bar{x}$	...	1,156	1,017	0,779

Cálculo de  $F$  para verificar la homogeneidad de las varianzas:

— prueba de Hartley:  $F = 2,202$  ( $p > .05$ ),

— prueba de Barlett:  $F = 1,156$  ( $p > .05$ ).

Prueba de «t» aplicada a la diferencia entre las medias:

— grupos 1.º y 2.º;  $t\bar{X}_1 - \bar{X}_2 = 1,669$  ( $p > .05$ );

— grupos 1.º y 3.º;  $t\bar{X}_1 - \bar{X}_3 = 0,269$  ( $p > .05$ );

— grupos 2.º y 3.º;  $t\bar{X}_2 - \bar{X}_3 = 0,098$  ( $p > .05$ ).

Por lo tanto, puede concluirse que, con respecto a los conocimientos previos, la composición de los grupos guarda la necesaria homogeneidad y que no debe esperarse que las diferencias existentes condicionen una diferencia de aprendizaje a lo largo de la experimentación.

TABLA 5

DISTRIBUCION DE LAS PUNTUACIONES DE ACTITUDES PREVIAS A LOS TRES GRUPOS

GRUPO 1		GRUPO 2		GRUPO 3	
Participante	Puntuación	Participante	Puntuación	Participante	Puntuación
A	1	B	3	C	5
F	3	E	2	D	3
G	2	H	3	I	2
L	1	K	2	J	3
M	3	N	2	O	3
R	5	Q	2	P	3
S	2	T	4	U	3
Y	4	X	2	Y	2

## ESTADÍSTICOS Y PARÁMETROS

N ... ..	8	8	8
$\bar{X}$ ... ..	2,625	2,500	3,000
S ... ..	1,317	0,707	0,866
S <sup>2</sup> ... ..	1,734	0,500	0,750
$\sigma$ ... ..	1,408	0,756	0,926
$\sigma_{\bar{x}}$ ... ..	0,500	0,267	0,327

Cálculo de  $F$  para verificar la homogeneidad de las varianzas:

— prueba de Hartley:  $F = 3,469$  ( $p > .05$ ),

— prueba de Barlett:  $F = 3,139$  ( $p > .05$ ).

Prueba de «t» aplicada a la diferencia entre las medias:

— grupos 1.º y 2.º;  $t\bar{X}_1 - \bar{X}_2 = 0,221$  ( $p > .05$ );

— grupos 1.º y 3.º;  $t\bar{X}_1 - \bar{X}_3 = 0,629$  ( $p > .05$ );

— grupos 2.º y 3.º;  $t\bar{X}_2 - \bar{X}_3 = 1,183$  ( $p > .05$ ).

En consecuencia, se puede afirmar que la composición de los tres grupos experimentales reúne la necesaria condición de la homogeneidad respecto de la actitud previa de sus integrantes hacia la estadística.

#### 4.2. Análisis de sus resultados a nivel de «postest».

TABLA 6

PUNTUACIONES DE CONOCIMIENTOS A NIVEL DE PRETEST  
Y DE POSTEST

PARTICIPANTE	«Pretest» (X)	«Postest» (Y)
A ... ..	12	17
B ... ..	11	18
C ... ..	9	14
D ... ..	8	12
E ... ..	8	16
F ... ..	6	8
G ... ..	6	8
H ... ..	6	13

PARTICIPANTES	«Pretest» (X)	«Postest» (X)
I ... ..	6	18
J ... ..	6	19
K ... ..	6	9
L ... ..	5	9
M ... ..	5	8
N ... ..	5	17
O ... ..	5	21
P ... ..	4	19
Q ... ..	4	14
R ... ..	3	14
S ... ..	3	14
T ... ..	3	7
U ... ..	3	10
V ... ..	3	13
X ... ..	2	4
Y ... ..	1	10

## ESTADÍSTICOS Y PARÁMETROS

N ... ..	24	24
$\bar{X}$ ... ..	5,417	12,875
S ... ..	2,644	4,456
S <sup>2</sup> ... ..	6,991	19,859
$\sigma$ ... ..	2,701	4,552
$\sigma\bar{x}$ ... ..	0,551	0,929
$r = 0,394 (p < .05)$		

Prueba de la homogeneidad de las varianzas (Hartley):

$$F = \frac{19,859}{7,944} = 2,841; p (< .05).$$

Prueba de Barlett:

$$F = 6,135 (< .05).$$

Prueba de «t» aplicada a la diferencia entre medias de muestras correlacionadas:

$$\begin{aligned} \sigma \text{ dif. } \bar{X} - \bar{Y} &= \\ &= \sqrt{(0,551^2 + 0,929^2) - 2 (0,394 \cdot 0,551 \cdot 0,929)} = 3,161; \\ t &= \frac{12,875 - 5,417}{3,161} = 8,533 (p < .001). \end{aligned}$$

En consecuencia, hay que rechazar al nivel de confianza del 999 por 1.000 la homogeneidad de las medias a nivel de *pretest* y de *postest*.

TABLA 7

PARTICIPANTE	«Pretest»	Grupo asignado	«Postest»
A	1	1	1
B	3	2	1
C	5	3	3
D	3	3	3
E	2	2	2
F	3	1	3
G	2	1	2
H	3	2	2
I	2	3	2
J	3	3	2
K	2	2	2
L	1	1	1
M	3	1	2
N	2	2	2
O	3	3	2
P	3	3	2
Q	2	2	2
R	5	1	5
S	2	1	2
T	4	2	3
U	3	3	3
V	2	3	2
X	2	2	1
Y	4	1	3

ESTADÍSTICOS Y PARÁMETROS

$N$	24	24
$\bar{X}$	2,708	2,208
$S$	1,020	0,865
$S^2$	1,040	0,748
$\sigma$	1,042	0,884
$\sigma\bar{x}$	0,213	0,180
$r = 0,777$ ( $p < .01$ )		

Pruebas de homogeneidad :

$$F \text{ (Hartley)} = \frac{1,040}{0,748} = 1,389 \text{ (} p > .05 \text{)}.$$

$$F \text{ (Barlett)} = 0,207 \text{ (} p > .05 \text{)}.$$

$t$  = (para muestras correlacionadas) :

$$\sigma \text{ dif. } \bar{X} - \bar{Y} =$$

$$= \sqrt{0,213^2 + 0,180^2 - 2(0,7774 \cdot 0,2126 - 0,1803)} = 0,1345$$

$$t = \frac{2,708 - 2,208}{0,1345} = 3,717 \text{ (} p < .01 \text{)}.$$

En consecuencia, pese a existir una diferencia significativa, no puede probarse que se haya producido un incremento significativo en la actitud hacia la estadística, ya que no se puede rechazar la hipótesis nula respecto a la homogeneidad de las varianzas.

Tras las dos pruebas anteriores, puede rechazarse la hipótesis nula a un elevadísimo nivel de confianza y afirmar, en consecuencia, que se ha producido, en los tres grupos experimentales, un incremento en el nivel de conocimientos y en la actitud positiva hacia la estadística.

Esta afirmación tiene un carácter genérico que debe verificarse para cada una de las fuentes de variación que han podido afectar a ambas variables dependientes —conocimientos y actitud— en cada uno de los tres grupos experimentales.

Para ello, hemos procedido a realizar el análisis de la varianza. Nótese que, al no existir, más que en un solo caso, una correlación significativa entre los resultados del *pretest* y los del *postest* en conocimientos y en actitudes, hemos rechazado la posibilidad de verificar el análisis de la covarianza.

Veamos los dos tratamientos por separado. En primer lugar, los cálculos referentes a los conocimientos.



TABLA 8

## DISTRIBUCION DE LOS RESULTADOS DE CONOCIMIENTOS (PRETEST Y POSTEST)

GRUPO 1			GRUPO 2			GRUPO 3		
PARTICIPANTE	«Pretest»	«Postest»	PARTICIPANTE	«Pretest»	«Postest»	PARTICIPANTE	«Pretest»	«Postest»
A ... ..	12	17	B ... ..	11	18	C ... ..	9	14
F ... ..	6	8	E ... ..	8	16	D ... ..	8	12
G ... ..	6	8	H ... ..	6	13	I ... ..	6	18
L ... ..	5	9	K ... ..	6	9	J ... ..	6	19
M ... ..	5	8	N ... ..	5	17	O ... ..	5	21
R ... ..	3	14	Q ... ..	4	14	P ... ..	4	19
S ... ..	3	14	T ... ..	3	7	U ... ..	3	10
Y ... ..	1	10	X ... ..	2	4	V ... ..	3	13
$r = 0,349 (p > .05)$			$r = 0,753 (p < .05)$			$r = 0,016 (p > .05)$		
$\Sigma X$ ... ..	41	88	$\Sigma X$ ... ..	45	98	$\Sigma X$ ... ..	44	126
$\Sigma X^2$ ... ..	285	1049	$\Sigma X^2$ ... ..	311	1380	$\Sigma X^2$ ... ..	276	2096
$(\Sigma X)^2$ ... ..	1681	7744	$(\Sigma X)^2$ ... ..	2025	9604	$(\Sigma X)^2$ ... ..	1936	15876

Resultados totales:  $\Sigma X = 442$ ;  $\Sigma X^2 = 5402$ ;  $(\Sigma X)^2 = 38866$ .

$$C = \frac{442^2}{48} = 4070,083.$$

$$\text{Suma de cuadrados intergrupo: } \frac{38866}{8} = 4858,250 - 4070,083 = 788,167.$$

Suma de cuadrados total:  $5402 - 4070,083 = 1331,917$ .

Suma de cuadrados intergrupos:  $1331,917 - 788,167 = 543,750$ .

FUENTE DE VARIACION	Suma de cuadrados	Grados de libertad	Media cuadrática	F	Nivel de significación
Intergrupo ... ..	788,167	5	157,633	12,176	( $p < .01$ )
Intragrupo ... ..	543,750	42	12,946		
TOTAL ... ..	1331,917	47			

Calculamos, seguidamente, la varianza debida a las distintas fuentes de variación intergrupo:

- a) evaluación previa-final (*pretest* y *posttest*),
- b) diferencia entre los tres métodos,
- c) interacción entre a) y b).

Es obvio que, al referirme a a), estoy aludiendo, además de a alguna variable independiente, a la maduración, historia personal y al *pretest*<sup>12</sup>. Por el contrario, cuando me refiero a la fuente b), estoy hablando de las diversas variables independientes cuyos efectos sobre los conocimientos y la actitud tratamos de apreciar.

a) *Evaluación previa-final:*

P R U E B A S			
	«Pretest»	«Postest»	
$\Sigma X$	130	312	$(\Sigma X)^2$ total:
$(\Sigma X)^2$	16900	97344	114244

Suma de cuadros correspondiente a la variación *pretest-postest*:

$$\frac{114244}{24} - 4070,083 = 690,084$$

<sup>12</sup> Cfr., más delante, el apartado «Evaluación del diseño factorial» (punto 5).

b) *Diferencia entre los tres métodos:*

	MÉTODOS			
	1.º	2.º	3.º	
$\Sigma X$	129	143	170	$(\Sigma X)^2$ total:
$(\Sigma X)^2$	16641	20449	28900	65990

Suma de cuadrados correspondiente a la diferencia entre los tres métodos:

$$\frac{65990}{16} = 4124,375 - 4070,083 = 54,292.$$

Prueba de «t» respecto de la homogeneidad de las medias de las muestras:

a) *Evaluación previa-final:*

Grupo 1.º:  $t = 3,466$  ( $p < .01$ ),

Grupo 2.º:  $t = 3,218$  ( $p < .01$ ),

Grupo 3.º:  $t = 6,359$  ( $p < .01$ ).

b) *Diferencia de los tres métodos a nivel de «postest».*

Prueba de la homogeneidad de las varianzas:

$$F \text{ (Barlett)} = 1,080 \text{ } (p > .05).$$

Prueba de la homogeneidad entre las medias de las tres muestras:

— Método 2.º - método 1.º:  $t = 0,574$  ( $p > .05$ ).

— Método 2.º - método 3.º:  $t = 1,535$  ( $p > .05$ ).

— Método 3.º - método 1.º:  $t = 2,534$  ( $p < .05$ ).

FUENTE DE VARIACION	Suma de cuadrados	Grados de libertad	Media cuadrática	F	Nivel de significación
a) Evaluación previa-final ... ..	690,084	1	690,084	53,305	( $p < .01$ )
b) Tres métodos ... ..	54,292	2	27,146	2,097	( $p > .05$ )
c) Interacción $a \times b$ ... ..	43,791	2	21,896	1,691	( $p > .05$ )
Intragrupo ... ..	543,750	42	12,946		
TOTAL ... ..	1331,917	47			

Concluimos, pues, que el incremento de aprendizaje se da en los tres grupos experimentales independientemente del método con el que haya trabajado. Por el contrario, no hemos encontrado una diferencia significativa entre unos y otros, ya que, aunque parece haberla entre los métodos tercero y primero, a favor de aquél, no podemos afirmar que las tres muestras tengan distintas varianzas.

*Análisis de la varianza de los resultados correspondientes a actitudes.*

En consecuencia, no procede rechazar la hipótesis nula respecto de la homogeneidad de las varianzas de los seis «grupos» (pruebas de *pretest* y *posttest* de los tres grupos). Por ello, no aportaré los cálculos de las pruebas de «t» respecto de las diferencias entre las medias.

Sobre esta base no podemos afirmar que se haya producido, en términos válidos para toda la población implicada en el experimento, un cambio positivo en la actitud susceptible de poder ser generalizado a otras poblaciones diferentes a ésta.

*Cálculo de la fiabilidad de la prueba.*

Independientemente del experimento se ha procedido a realizar este cálculo. Ello se ha efectuado por el método de mitades (cuestiones pares e impares), aplicando, luego, la fórmula de Spearman-Brown. Dado que la prueba de conocimientos ocupa 24 de las 25 cuestiones se ha prescindido de las respuestas aportadas a la última —a la pregunta sobre actitudes— para equilibrar el número de cuestiones pares e impares.

Resultados:

- $r$  en la prueba previa:  $r = 0,522$  ( $p < .01$ ),
- $r$  en la prueba final:  $r = 0,712$  ( $p < .01$ ),
- coeficientes de fiabilidad respectivos:  
0,686 ( $p < .01$ ) y 0,823 ( $p < .01$ ).



TABLA 9

## DISTRIBUCION DE LOS RESULTADOS DE ACTITUDES (PRETEST Y POSTEST)

GRUPO 1			GRUPO 2			GRUPO 3		
PARTICIPANTE	«Pretest»	«Postest»	PARTICIPANTE	«Pretest»	«Postest»	PARTICIPANTE	«Pretest»	«Postest»
A ... ..	1	1	B ... ..	3	1	C ... ..	5	3
F ... ..	3	3	E ... ..	2	2	D ... ..	3	3
G ... ..	2	2	H ... ..	3	2	I ... ..	2	2
L ... ..	1	1	K ... ..	2	2	J ... ..	3	2
M ... ..	3	2	N ... ..	2	2	O ... ..	3	2
R ... ..	5	5	Q ... ..	2	2	P ... ..	3	2
S ... ..	2	2	T ... ..	4	3	U ... ..	3	3
X ... ..	4	3	Y ... ..	2	1	V ... ..	2	2

$$r = 0,94 \ (p < .01)$$

$$r = 0,48 \ (p > .05)$$

$$r = 0,62 \ (p > .05)$$

$\Sigma X$ ... ..	21	19	$\Sigma X$ ... ..	20	15	$\Sigma X$ ... ..	24	19
$\Sigma X^2$ ... ..	69	57	$\Sigma X^2$ ... ..	54	31	$\Sigma X^2$ ... ..	78	47
$(\Sigma X)^2$ ... ..	441	361	$(\Sigma X)^2$ ... ..	400	225	$(\Sigma X)^2$ ... ..	576	361

Sumas totales:  $\Sigma X = 118$ ;  $\Sigma X^2 = 336$ ;  $(\Sigma X)^2 = 2364$ .

$$C = \frac{1182}{48} = 290,083.$$

$$\text{Suma de cuadrados intergrupo} = \frac{2364}{8} = 295,5 - 290,083 = 5,417.$$

$$\text{Suma de cuadrados total} = 336 - 290,083 = 45,917.$$

$$\text{Suma de cuadrados intragrupo} = 45,917 - 5,417 = 40,500.$$

FUENTE DE VARIACION	<i>Suma de cuadrados</i>	<i>Grados de libertad</i>	<i>Media cuadrática</i>	F	<i>Nivel de significación</i>
Intergrupo ... ..	5,417	5	1,083	1,123	(p > .05)
Intragrupo ... ..	40,500	42	0,964		
TOTAL ... ..	45,917	47			

*Evaluación del diseño factorial.* Respecto de los objetivos iniciales se puede concluir lo siguiente:

1. Se ha conseguido un incremento estadísticamente significativo respecto de los conocimientos de los participantes sobre estadística descriptiva y muestras, entre el *pretest* y el *posttest*.
2. Se ha logrado un incremento en el interés respecto de este tema, aunque no tan elevado como para poder generalizarlo a nivel de «universo».
3. No se puede dar por probada —de una manera estadísticamente significativa— la diferencia entre los tres métodos de trabajo experimentados: 1, lección magistral; 2, lectura de un documento técnico seguida de lección magistral; 3, lectura de un documento técnico seguida de una discusión sobre el tema. No obstante, los datos prueban que, a nivel de *estos* grupos, sí se ha operado un incremento mayor de aprendizaje con el método tercero que con el segundo, y con éste respecto del primero.
4. El que no se hayan encontrado diferencias estadísticamente significativas entre los tres métodos, pese a la diferencia de «coste» de uno a otro <sup>13</sup>, nos debe animar a no minusvalorar la importancia didáctica de la lección magistral, particularmente cuando se le emplea de forma asociada con un documento escrito sobre el mismo tema. Es posible que una motivación adecuada —como la que ha podido provocar la situación experimental— unida a un método aparentemente desventajoso llegara —en un *infinito* número de

---

<sup>13</sup> Recuérdense los tres métodos: 1.º, lección magistral; 2.º, lectura de un documento seguida de una lección magistral; 3.º, lectura de un documento seguida de discusión con aclaración de dudas. Nótese, además, que el primer método implicaba un considerable ahorro de tiempo (aproximadamente del 40 por 100). No hemos utilizado ningún grupo de control dada la situación de trabajo en el que lógicamente se debían situar todos los participantes en el programa.

aplicaciones— a provocar un aprendizaje no inferior al de otros métodos más complejos.

5. El diseño con el que aquí se ha trabajado

O	M	X	O	O = <i>pretest</i> y <i>postest</i>
O	M	X'	O	M = muestreo estratificado
O	M	X''	O	X, X', X'' = tres variables independientes

... nos permite asegurar que se han controlado las posibles fuentes de invalidez interna de <sup>14</sup>:

- historia (dado el escaso lapso de tiempo transcurrido entre el *pretest* y el *postest*);
- maduración (ídem);
- instrumentación (por haber sido la misma prueba la aplicada en el *pretest* y en el *postest*);
- selección (dado el criterio utilizado para la adscripción de los participantes a los grupos, criterio superior al muestreo aleatorio);
- mortalidad (no ha habido ninguna);
- interacción de maduración y selección, etc.

Por el contrario, no podemos afirmar que se hayan controlado suficientemente los factores de invalidación de:

- realización de la prueba (por ser *pretest* = *postest*),
- regresión (dado que sólo cuatro de los ocho coeficientes de correlación entre los resultados del *pretest* y los del

<sup>14</sup> Utilizo los criterios citados por D. CAMPBELL y J. STANLEY, *op. cit.*, páginas 16-38.

*postest* son significativos para  $p < .05$ ). No obstante, debe tenerse presente que no está afectada por el criterio de selección.

#### COMENTARIO FINAL

Los dos ejemplos presentados aquí representan acaso los extremos entre los cuales puede situarse el tipo de investigación a realizar por los profesores en ejercicio. El primero es muy elemental, pero supone un primer paso a dar. El segundo es bastante más complejo, ya que exige el dominio de algunos procedimientos estadísticos precisos, por lo que se encuadra mejor dentro de la investigación aplicada que en el de la investigación activa.

En el cuadro real de las posibilidades de perfeccionamiento de los profesores dentro y fuera de sus centros educativos, tal tipo de investigación debe desarrollarse siempre en el marco de un trabajo en equipo y con el asesoramiento de un profesor —de ciencias, por ejemplo— o de un pedagogo o psicólogo respecto de los procedimientos experimentales.

Otra posibilidad la constituye la realizada por nosotros: el asesoramiento desde un Instituto de Ciencias de la Educación a varios profesores o equipos de profesores en la realización de proyectos de investigación activa.

Con todo ello se puede ayudar a cubrir el sentido propio de la investigación en el trabajo educativo. La etimología del término investigación —o de su sinónimo «indagación»— apunta hacia una meta a la que deben llegar los profesores, *rastrear, seguir el rastro o la pista de algo*. Según esto, el profesor investiga cuando sigue los pasos de su propio trabajo.